

# Introduction

# Introduction

---

The Office of Technology Assessment (OTA) has had a longstanding interest in the use of randomized clinical trials (RCTS). The OTA report *Assessing the Efficacy and Safety of Medical Technologies* (225) discusses the advantages and disadvantages of RCTS and puts forward a number of policy-alternatives for identifying technologies in need of assessment, stimulating clinical trials, and disseminating information derived from them. *The Implications of Cost-Effectiveness Analysis of Medical Technology* (229) discusses the value of RCTS in cost-effectiveness analyses, and notes that information derived from RCTS is not available on many technologies. *Strategies for Medical Technology Assessment* (234) concludes that RCTS are the “definitive experimental method for evaluating the efficacy or health benefits of a technology.” Other OTA assessments and case studies in some way use or discuss the results of RCTS (e.g., case studies for *The Implications of Cost-Effectiveness Analysis of Medical Technology*, 1978-1982; *A Review of Selected Federal Vaccine and Immunization Policies*, 1979; *Technology Transfer at the National Institutes of Health*, 1982; *Post-marketing Surveillance of Prescription Drugs*, 1982).

OTA’S continuing interest in RCTS led to the question that this study posed: What has been the impact of RCTS on health policy and medical practice? This study is based largely on a review of the literature concerning the history of RCTS and their support, their use in health policymaking, and their influence on medical practice. This review has been supplemented by discussions with policymakers and medical and health specialists with particular interests in RCTS.

The remainder of this chapter contains background material about RCTS and a brief discussion of the diffusion of medical technologies. Chapter 2 covers the funding of RCTS and some nonrandomized clinical trials. The current and possible future uses of RCTS in health policymaking are discussed in chapter 3. Chapter 4 looks at criticisms of and alternatives to RCTS, and the

characteristics of RCTS that appear to influence their impact. Chapters reviews the literature specifically about the impacts of RCTS on medical practice. Suggestions for strengthening the impact of RCTS are brought together in the last chapter.

In this paper, “medical technologies” include drugs, devices, and medical and surgical procedures. The organizational and supportive systems through which medical care is provided are part of medical technology in its broadest sense, but they are not discussed here in detail.

Drugs, devices, and procedures are used to diagnose, treat, and prevent disease, and to promote health. Diagnosis usually involves tests and procedures, often using specific medical devices. Treatments may include the use of drugs, devices, and procedures. Disease prevention is traditionally broken down into the categories of primary, secondary, and tertiary prevention. Primary prevention is aimed at avoiding disease altogether. Most vaccines, for instance, are considered primary prevention. Secondary prevention consists of strategies to detect disease in its early stages of development, with the hope of improving patient outcome. Many screening programs, e.g., for breast cancer, are examples of secondary prevention. Tertiary prevention attempts to arrest further deterioration in individuals who suffer later stages of disease. RCTS can be used in evaluations of all types of disease prevention.

RCTS are experiments that test the safety and efficacy of medical technologies. An “experiment” more generally has been defined as “[t]he planned manipulation of material, subjects, or processes by the experimenter, in order to establish a cause-effect relation or a rule (model) for the variation of observations” (151).

In this century, RCTS have replaced anecdotal evidence as the standard for evaluating medical technologies. The development and increasing use of RCTS in evaluating medical interventions is not an isolated phenomenon, but rather part of a broader trend. Experimental methods are increas-

ingly used in studying all types of human problems. In or out of the clinical setting, the randomized trial is the strongest tool available across a spectrum of research topics (56,198). For example, the testing and evaluation of social interventions using randomized designs forms the basis for the growing field of social experimentation. Social and medical issues meet in health services research in evaluating interventions that are not medical technologies, but that are applied in clinical settings. For example, in an innovative program at Cleveland Metropolitan General Hospital researchers have conducted randomized trials on the effect on physicians' ordering of tests when they are provided information or education (168). McGhan and colleagues (148) report a randomized trial comparing pharmacists and technicians as dispensers of prescriptions for ambulatory patients. The use of randomized trials in this field will undoubtedly grow, as it could greatly contribute to the efficient provision of health services. While the study designs in this field are identical or similar to those used to test medical technologies, these studies will not be discussed in detail in this paper.

In clinical settings, RCTS occupy a niche at one end of the spectrum of biomedical research. At the other is found untargeted basic research in biological processes, moving toward preclinical and clinical research and the development of medical technologies for specific diseases. The RCT is a method for testing the efficacy and safety of such technologies. The reason for conducting an RCT should be a sound hypothesis about the technology in question. Fisher (73) notes that the significance of preclinical laboratory research and of clinical trials in fact depend on each other:

Until a proper clinical test is carried out, no matter how promising a line of investigation seems to be it remains just that, a promise. Clinical research, on the other hand, without a firm biological basis acquired from laboratory investigation is apt to be nothing more than product testing.

Like other kinds of experiments, the RCT compares the effect of an intervention (a medical technology) on one group of people with the fate of

a "control" group, which is not subject to the intervention but is otherwise similar to the "experimental" group. RCTS are distinguished from other kinds of comparative studies in that individuals are randomly assigned to these different groups. "Random" does not mean "haphazard" in this case, but rather that individuals are assigned with equal probability to the experimental or the control group.

Randomization is crucial in allowing certain statistical inferences about the experiment's outcome. Random allocation eliminates overt and covert biases in the assignment of patients to treatments. Patients with particular medical characteristics are not determinedly placed more frequently in any one group. Differences in the outcomes of the groups can thus be attributed to the intervention, within the limits of statistical probability.

In other comparative studies, groups are formed by methods other than randomization. But experimenters may be biased in selecting the members of these groups because, consciously or unconsciously, they favor some particular outcome. Such bias would of course compromise the conclusions about why any difference is observed between the groups. Other kinds of epidemiologic and evaluative studies can provide valuable information, though they cannot replace RCTS. See *Strategies for Medical Technology Assessment* (234) for information about the role of other study designs in assessing medical technologies.

The design and execution of RCTS may benefit from prior nonrandomized clinical studies, such as case reports and retrospective analyses of clinic records. "Suggestive evidence" from these sources may provide the justification for carrying out an RCT, and indicate patients most likely to benefit from the technology. The suggestive evidence that "lumpectomy" (removing only a tumor and small amount of tissue) might be effective in treating breast cancer came from retrospective examination of clinic records. An RCT based on that evidence confirmed the value of lumpectomy (188).

Further details about the rationale and methods of RCTS are described in later sections of this chapter.

## BACKGROUND

For as long as medical care has been given, people have been concerned about its effects. Does a given treatment cure, prevent, or ameliorate a condition, and what are its other effects, beneficial and detrimental? Nevertheless, specific questions about a treatment's efficacy and safety, not to mention cost effectiveness, have not always been explicit, and attempts to answer them even less so. Concern about the effects of medical interventions has been heightened by three developments in recent decades: the development of more powerful medical technologies; the availability of more effective tools to evaluate them, e.g., the RCT; and the rapidly increasing costs of health care,

During the latter half of the 19th century, quantitative evaluation led to the abandonment of a substantial number of therapies, with no effective therapies to replace them. Major breakthroughs in medical treatment and disease prevention began in the late 19th century and continued through the 1930's and 1940's, brought about by greater understanding of infectious diseases. The advances were obvious, and confidence in medicine ran high. The successes in overcoming many infectious diseases made chronic diseases the major causes of sickness and death in developed countries, and led to new kinds of medical interventions. As success stories became fewer and less dramatic, uncertainty arose again about the value of medical practices.

The rising cost of medical care is one of the most pervasive issues in health care. The development and analysis of strategies to control costs is an

area of research itself (see, e.g., 230). New technologies in particular contribute to the rise in both capital costs (e. g., for the new generation of diagnostic imaging equipment such as computed tomography scanners and nuclear magnetic resonance imagers) and health manpower costs (e. g., for intensive care units and complex surgical procedures). Another fact of economic importance is that many technologies can be widely disseminated and used. Imaging, for example, is important in a wide range of medical practice, and new treatments for heart disease address the most frequent chronic disease and cause of death in this country.

The combined concerns for the safety and efficacy of medical practices and for the rising costs of health care together impel the need for rational decisionmaking to avoid what does not work or is unsafe and to get the most for health care dollars. Such decisionmaking depends on information that compares the safety and efficacy of competing technologies. The best method of gathering such information is the RCT.

It has been estimated that between 10 and 20 percent of all current medical procedures have been shown efficacious in controlled trials (225). While it is not possible or desirable to evaluate all medical practices with RCTs, the method could be used much more in evaluating new technologies, in evaluating new applications of existing technologies and in evaluating practices that have long been used but that are still of questionable value (e.g., hysterectomy for some indications).

## A BRIEF HISTORY OF THE RCT IN MEDICINE

RCTs are a product of this century, but their forerunners in evaluating "health technologies" reach back at least to Biblical times and in all probability much earlier. An essential element of RCTs, the use of a control group, is related in the Book of Daniel (ch. 1). Daniel was among those children of Israel "in whom was no blemish, but well favored, handsome and skillful in all wis-

dom" who were chosen to be readied to serve Nebuchadnezzar, the conquering king. Placed in the charge of the prince of the eunuchs, the children were to be fed the king's meat *and* wine. Daniel, not wanting to be defiled by the diet, asked of the eunuch that he and his three companions from Judah be given pulse (a type of pea) and water instead. The eunuch was afraid he would

be blamed for the poor condition of the boys that he thought would certainly result from such nutrition. Daniel convinced him to give them pulse and water for 10 days:

Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat.

Ten days later Daniel and his companions were judged "fairer and fatter in flesh" than the other children, and the impact of the trial was immediate and direct. From then on, all the children were nourished on pulse and water.

Careful observation and the use of comparison groups have marked advances for human well-being since Daniel's time. Only careful evaluation satisfies healthy scientific skepticism about the value of new technologies. Unfortunately, the need for experimentation is not universally acknowledged, and there are undoubtedly those in medicine today who subscribe to an updated version of the reasoning of a respected 19th century physician: given irrefutable evidence that blood circulates, he replied: "Experiments irritate nature. When nature is irritated it acts otherwise than when it is left alone. Therefore, experiments prove nothing" (94). Nonetheless, progress has been made.

James Lind, in his famous 1747 experiment, compared six treatments for the prevention of scurvy. A full 150 years after the treatment was first suggested in print, he confirmed citrus fruits as a successful prophylaxis (136). The impact of the trial was further delayed: it was 40 years before the British navy required that citrus fruits be carried on ships at sea (40).

The tradition of careful observation and comparison was joined in this century with quantitative methods, to produce modern experimental design (151). In the 1920's and 1930's, R. A. Fisher developed methods for statistical inference based on random allocation, which he applied to his agricultural experiments. Fisher led the way for the medical application of randomization and the statistical methods reliant on random allocation.

The value of knowing which was in fact the first "true RCT" is debatable, but the history is interesting. A. B. Hill was the first major advocate for

RCTS in England, where he carried out a trial of patulin against the common cold in 1944 (175) and a trial of streptomycin therapy for tuberculosis, begun in 1946 (161). W.G. Cochran was the earliest strong proponent of RCTS in this country. Some contend that a trial of therapy for tuberculosis published in 1931 by Amberson and colleagues (1) qualifies as the first RCT. In their trial, the control and treatment groups were closely matched on various clinical dimensions, with the choice of which group would get the experimental treatment decided by the flip of a coin. They clearly recognized the value of unbiased allocation, but not the importance of randomization for valid statistical evaluations. Hill, on the other hand, clearly had emphasized randomization (141). Whether Amberson or Hill conducted the "first RCT" is thus a question of whether the experimenter's full awareness of its principles are included in the definition.

The present study concerns the modern RCT, which began with the randomized allocation to treatment groups in clinical settings. This procedure was introduced around the middle of this century at about the same time as the modern generation of drugs, including antibiotics, and vitamins, and other therapeutic measures were developed, demanding standards for evaluation. Adopted initially to evaluate drugs and vaccines, the RCT still enjoys its widest use in that area, its use in evaluating medical procedures and devices developing later and more slowly. The move from using the RCT in evaluating therapies and preventive interventions for acute diseases, to its use in treating and preventing chronic diseases occurred first during the late 1950's in tests of new treatment regimens for leukemia. In the 1960's, RCTS were employed in developing treatment regimens for other chronic diseases, notably cardiovascular diseases. They have also been used in testing diagnostic techniques (e.g., mammography to detect breast cancer), though still infrequently.

The use of RCTS has shown steady growth. In a random sample of articles from general medical journals, no-RCTs were reported in 1946, while 5 percent were reports of RCTS in 1976 (75). In an exhaustive search of the literature in English through 1981, Haines (103) found 51 RCTS related to neurosurgery; half of those had been published

since 1977. The growing number of RCTS and interest in them resulted in adding the heading “clinical trials” to *Index Medicus* in 1980. Another sign

of this growing interest was the founding of the Society for Clinical Trials in 1978 to encourage exchange about methodological issues (see box A).

#### Box A.—The Society for Clinical Trials

The Society for Clinical Trials was founded in 1978 by a group of individuals with experience in clinical testing, epidemiology, statistics, and computer science. It was formed to allow greater exchange about methodological issues and about the impacts of RCTS, topics that are rarely addressed in medical periodicals, even in reports of trials.

The Society has more than 1,000 members. It sponsors an annual meeting and publishes the quarterly journal *Controlled Clinical Trials*. Its main objective is “to promote the development and exchange of information for design and conduct of clinical trials and research using similar methods.” The society’s specific long-term objectives include the following (209):

- Promotion of methodological research emphasizing design, organization, operation, and analysis.
- Promotion of the application of sound principles to design, operation through workshops and meetings sponsored by the organization. Some of these workshops and meetings maybe international in character and held in countries other than the United States.
- Promotion of better communication by development, where possible, of standard terminology.
- Promotion of better understanding to those entering this field by serving as an important resource for the design and conduct for these studies.
- Promotion of better communication through the development of standards for the analysis and reporting of results.
- Promotion of better understanding by the general public of the importance of clinical trials for the evaluation of health care procedures.

## A Description of the Method

### General Structure

Fisher’s rationale for randomizing as a valid basis for statistical inference is still the touchstone of RCT methodology. RCTS are actually a family of study designs that share the feature of randomized assignment to treatment groups.

In the simplest of these designs, individuals with a condition in common (e. g., the common cold) are allocated to two groups by an accepted randomization procedure (e.g., using random number tables or computer-generated random numbers). A promising but unproven technology (e.g., a *new* drug) is applied to one group, while the other is given the standard treatment, if one exists. The control group may be given no treatment at all, if that is standard, or preferably, when possible, a placebo that resembles the experimental drug. At an appropriate time after applying the

technology each individual in the two groups is **assessed** for a prespecified outcome. The outcome can be death or a signal health event (e.g., a heart attack) or an intermediate physiological measure, such as a change in blood pressure. In a vaccine trial and some drug trials, presence or absence of disease after some time is an appropriate endpoint. The aggregate results for each group are then compared. Statistical tests are applied to the results to determine whether or not the new technology is better than the old.

In a well-designed trial, both the numbers of participants and the endpoints are chosen so that there is a reasonable probability that a statistically significant result can be obtained, if in fact the treatments being compared differ by some prespecified amount or more. While simple in theory, in practice RCTS are complex undertakings. Klimt (123) describes five phases in RCTS:

1. A Planning Phase that precedes general funding.
2. After approval of a broad outline of design and funding a Preparatory Phase, the protocol, the forms, and the organization are laid down in detail,
3. The Recruitment Phase that starts with the acquisition of the required number of clinical units and is followed by the recruitment of patients.
4. The Follow-up and Termination Phase during which no further recruitment takes place but patients are followed for the requisite number of years. The length of follow-up is determined by the nature of the disease and the kind of treatment effect expected. The termination part of this phase requires clean-up of the data base on patient information collected and final classification of endpoints.
5. Last, the Analysis Phase, where no new data are being gathered, the statistical analysis is performed, conclusions are drawn, and papers written,

Each phase presents its own challenges. The practical problems and basic guidance are discussed in the journal literature and in a limited number of texts, for example *Fundamentals of Clinical Trials*, by Friedman, Furberg, and DeMets (84), is an excellent reference. In addition, Peto and colleagues (180,181) provide a detailed description of RCTS for the nonstatistician, including both their design and analytic features.

The size and complexity of RCTS vary greatly. Small-scale pilot studies with only a handful of patients may be undertaken by a single researcher. At the other extreme, thousands of patients in centers around the world may be participants in a single trial. Many of the recent RCTS supported by the National Heart, Lung, and Blood Institute, particularly those in primary and secondary prevention of cardiovascular disease, are large multicenter endeavors. For example, the recently completed Multiple Risk Factor Intervention Trial randomized 12,866 men at 22 clinical centers to test the effect of a multifactor intervention program on mortality from coronary heart disease (166).

Although all well-designed RCTS require a great deal of effort and thought in design and execution, multicenter trials present greater practical

problems. Well-conducted multicenter RCTS are characterized by such features as a centralized data collection center, a data monitoring committee (often of individuals independent of the study, with no vested interest in the trial or the intervention), and formal auditing procedures.

## Blinding

Because of bias for or against a treatment on the part of researchers and patients, and to control for the effect of expectations of outcome, (a natural human characteristic), the element of "blinding" also has become a characteristic of RCTS. The object of blinding is to prevent the awareness of which treatment is administered. When only the patient is unaware of the treatment the study is "single-blind;" when both the person administering treatment and the patient are unaware, it is "double-blind." Additional layers of blinding can be added. Often a person other than the treating physician evaluates patient outcome. That person can in turn be unaware of which group a patient is in. The statistician analyzing the data may do so blinded.

The most valuable tool for achieving blinding is use of a placebo, an inactive substance or procedure that mimics the intervention tested, so that those who are to be kept blind cannot tell it from the active intervention. Placebos are most often used in drug trials, though at least one surgical RCT, assessing internal mammary artery ligation for coronary artery disease, used a sham operation as a placebo for the control group. That practice would not be acceptable today, since even a sham operation, involving anesthesia and operative incision involves risk. Ethical placebos can be developed for some procedures, however. A recent RCT of apheresis for schizophrenia used sham pheresis in the controls (see ch. 5). In some cases blinding is clearly impossible, as in comparing a surgical with a medical procedure, or when patients and physicians can identify a given treatment because of its special side effects. If blinding is not possible, the effect of bias in unblinded studies can be minimized to the extent outcomes are measured by objective standards. Whatever the outcomes measured, even with no blinding, randomized allocation will lead to more reliable results than any other type of allocation.

## Techniques for Randomized Patient Allocation

Early randomization schemes were based on simple systems, such as the flip of a coin, alternate assignments of patients to groups as they arrived, or according to the day of the week they arrived, their birth dates, or their hospital or social security numbers.

Such methods have been abandoned for the most part, largely because the predictability of assignment allowed researchers and patients to manipulate assignments, or to selectively decide whether or not to participate in the trial. Assignments today are most often based on random number tables or computer-generated random numbers. Treatments may be assigned using pre-sealed envelopes, opaque to the light. In multi-center trials, assignments are often computer-generated by a central office when a participant is enrolled, and given to the physician over the telephone, allowing little scope for physician bias in assigning treatments.

In theory, randomization of all individuals into requisite groups for a trial cannot be improved on. Given a large enough sample size, factors affecting outcome will be distributed more or less equally among the groups. Logically, for smaller numbers of people, randomization produces greater equality among groups the more homogeneous the population, and the fewer the *prognostic factors* that affect the outcome. In practice, because patients and resources are not unlimited, and often patient populations are rather heterogeneous, techniques have been developed to improve the distribution of the number of patients and their prognostic factors among groups.

The chance imbalance of numbers of individuals in the groups can be prevented by a special

procedure called “random block permutation.” In effect, this technique ensures that after some pre-specified number of patients are entered in the trial, equal numbers are assigned to treatment groups.

“Stratification” is another commonly used, but controversial, method to better distribute factors of known prognostic importance during patient allocation. As individuals have entered the trial, they are classified by these factors, e.g., age, sex, and often diagnostic characteristics, e.g., extent of spread of a cancer. Randomization then takes place within these “strata,” that is within these particular subgroups.

The value of stratification in patient allocation is not uniformly agreed on (137), but stratification in analysis is a generally accepted procedure. In the latter, adjustments are made after the data have been collected to adjust for chance imbalance in prognostic factors between groups.

“Minimization” is a more recent idea for patient allocation (218). The technique takes into account a number of variables of prognostic interest, up to 15 or more, without forming mutually exclusive subgroups. As each participant is entered, a series of calculations is made to determine which assignment would minimize the differences between the groups. Different weights can be assigned to different patient variables according to their prognostic importance. If all are given equal weight, group assignments are made simply to distribute equally the largest number of variables. Randomized allocation is used only in assigning the first patient and when there is a “tie” and the same difference between groups would occur regardless of assignment. Minimization has become popular particularly in cancer trials, where a large number of factors are known to have prognostic importance (184),

## THE USES OF RCTS

The RCT was developed to discriminate between effective and ineffective treatments, particularly when the differences between treatments are moderate. More specifically, RCTS are used to accomplish the following:

- to compare the safety and efficacy of a new technology with a standard treatment, whether this is no treatment at all or a competing technology;
- to test the relative efficacy of a new technol-

ogy, assuming it has some other advantage over the standard, e.g., fewer side effects, lower cost;

- to determine the optimal way to use a technology to achieve a therapeutic effect; and
- to demonstrate the likely range of a technology's effectiveness in general practice as opposed to in highly controlled experimental settings. In a broader sense, RCTS can be used to answer questions susceptible to the scientific method about interventions involving human beings. Well-designed and executed RCTS are not merely product testing, but should answer questions about important hypotheses. They should, therefore, generate biologically and medically important information.

The results of RCTS may have widespread impact (143) insofar as they are used to allocate medical resources more efficiently (19,50,57,79,110,143); to effect the adoption and use of medical innovations (70,89,91,113,143); to hasten the abandonment of ineffective therapies (11,111); and to resolve controversies about competing treatments (170).

RCTS are most useful when either the benefit of a new treatment is uncertain or the relative ben-

efits of existing therapies are disputed (32). Thus, not all technologies need be evaluated in an RCT. Medical breakthroughs, such as the discovery of treatments like quinine for malaria, sulfa drugs and penicillin for bacterial infections, and insulin for diabetic acidosis, required no RCTS to demonstrate their efficacy. Startling breakthroughs, unfortunately, do not characterize most medical advances. Even in the case of breakthroughs, however, RCTS are useful to determine optimal treatment regimens. The current successful chemotherapy for Hodgkins disease was built up with stepwise RCTS after an initial breakthrough. Aside from breakthroughs, there are other technologies of accepted value that do not require the blessing of an RCT. For example (225):

. . . cast application for forearm fracture is a technology whose efficacy has been established by experience in medical settings. It illustrates a technology whose efficacy could be called "manifest," that is, whose efficacy and safety are obvious to the observer. Although alternatives to cast application might be as efficacious, its widespread acceptance in this country makes development and testing of other methods unlikely and probably unnecessary.

## THE ROLE OF THE PHYSICIAN IN RCTS

Traditionally, the physician has been the arbiter and judge of medical practices. It was presumed that careful observation of patients and reasoning about cause and effect would make the physician the best instrument to judge the success or failure of clinical practices. Until nearly the middle of this century, that presumption was largely unquestioned. Before the emergence of RCTS physicians were the only major actors in clinical decisionmaking. The growing importance of statistical evidence, and perhaps the growing importance of the statistician, was and is seen by some physicians as a threat. Some believe this response of physicians is a major impediment to the acceptance and adoption of good RCT results by the medical community (142):

To some extent the clinician's marginalization was implicit in the rationale for the RCT. Not only was the RCT viewed as capable of making finer, more reliable discriminations between the relative merits of effective therapies (112), but randomization was introduced because of its superiority over the clinical investigator in controlling for the variables which might affect therapeutic outcomes. Moreover, early critics of randomization have noted, the goal of minimizing the investigator's interpretive role is implicit in the logic of statistical hypothesis testing.

The extent to which physicians' feelings of displacement have affected the development and impact of RCTS is impossible to assess. It can now be judged only by anecdotal evidence, precisely

the standard that supporters of RCTS seek to replace. A more basic question than the one directly addressing RCTS may be a question about the role of research in general in clinical decisionmaking. Finally, it is important to understand the other factors that affect the way physicians treat patients.

Spodick (210) cites five behavioral pitfalls of physicians which affect both the conduct of RCTS and the acceptance of their results. The first is that the general acceptance of a practice is often taken for a proof of its effectiveness. The long use of bleeding, purging, and trephining provide examples. The rejection of “general acceptance” of a practice as adequate evidence of its efficacy underlay the 1962 amendments to the Food, Drug, and Cosmetics Act, which required “adequate and well-controlled studies” in support of new drug applications. Another pitfall of physician behavior is zeal, leading to glowing reports of success in the early applications of new practices. Such enthusiasm may be “inversely proportional to the quality of control” for treatments later shown ineffective or harmful in appropriately designed trials. Estrogen therapy for prostatic carcinoma, Vineberg implants for coronary artery disease, diethylstilbestrol to prevent spontaneous abortion, prophylactic portacaval shunts for portal hypertension, and internal mammary ligation are all practices that were enthusiastically embraced and have since been discarded because they lack efficacy or are unsafe,

A third pitfall is physicians’ uncritical acceptance of poor data. Poor data are often given as much credence as good, and more if they support a preconceived notion of what is right. Often, because the sheer volume of poor data is so great, small amounts of good data are not visible. Long before diethylstilbestrol was known to be harmful to women who were exposed before birth, six well-controlled trials had shown that the drug was ineffective. Seven other uncontrolled or poorly controlled trials had taken precedence while **50,000** pregnant women per year took the drug. A fourth related pitfall is blindness to what data exist.

The final pitfall is the “it can’t hurt mentality.” Even when practices are proven ineffective

through well-designed studies, they may still be continued. In some cases, no alternative treatment is available, and the physician feels that any treatment, even an ineffective one, is better than none. The physician may not always be wrong if “ineffective” is interpreted to include exploiting a placebo effect, or diverting patients from really harmful treatments. Unfortunately, however, there is never perfect knowledge about the effects of drugs or practices, and sometimes they may well “hurt” in the long term. The case of diethylstilbestrol illustrates this, as does the continued adherence to prescribing a bland diet, including cream, for peptic ulcer. There is some reason to believe that heavy intake of cream caused or accelerated atherosclerosis in some ulcer patients (40).

Spodick also speculates about the behavioral deterrents to initiating trials when they may be needed. Reverence for authority may cause physicians to adopt practices uncritically, i.e., when the practices are developed by and advocated by persons of renown. This was a factor in the widespread adoption of gastric freezing in treating peptic ulcer. Reverence for tradition makes it difficult to abandon an old practice, particularly when there is none to replace it. Physicians often feel a compulsion to treat, coupled with a reluctance to admit doubt. These attributes are often encouraged by patients. Physicians are also often loath to substitute clinical trial results for personal judgment in prescribing treatment. They may fear either withholding a new treatment or exposing patients to it, and therefore may be reluctant to participate in an RCT.

These *views* represent a fairly negative perception of physicians in relation to RCTS. On the positive side, it is physicians who initiate and participate in RCTS, and who form the majority of the method’s proponents. As in most fields, acceptance of new methods is bound to be gradual, partly owing to appropriate skepticism. The use and impact of RCTS has grown since the 1940’s, and the method itself is still evolving. Physicians and statisticians together are responsible for this progress, and there is evidence that physicians, including those in the community, are increasingly willing to participate in RCTS (see e.g., 65).

## THE DIFFUSION OF MEDICAL TECHNOLOGY\*

While it is useful to examine the effects of RCTS on the practice of medicine, it is useful to do so in the context of the larger questions of the adoption and use of medical technologies and the way medical practice changes.

The process by which a technology becomes part of the health care system is known as diffusion. Diffusion has two phases: the period when the decision is made to adopt the innovation, and the later period when decisions are made to use it. Research has focused on the first phase, as have Government policies. The use of a technology may be only tenuously related to its adoption. Each is discussed here in a separate section.

### The Adoption of Technologies

The adoption of technological innovations has captured the attention of hundreds of researchers, resulting in thousands of articles and many theories (72). Early research grew out of sociology (192), but much recent work has been done by economists (195). A tacit assumption in much of this research is that adopting an innovation is desirable.

The classical model describing diffusion of technology is an S-shaped curve, based on the concept of "contagion" or "spread" (72). The diffusion of technologies such as intensive care units and cardiac pacemakers has followed this pattern (195,227). At least one other model, the "desperation-reaction model," has been described by Warner (246). A first phase of explosive diffusion occurs because of a provider's sense of responsibility to the patient and their mutual desperation faced with a life-threatening situation. These responses are related to what Fox (76) has called "scientific magic," which is partly the tendency of medical practitioners to favor vigorous treatments and to be staunchly hopeful even when a positive outcome is unlikely. Cancer therapies often fit the desperation-reaction model: there are few effective tools to fight the disease, and little time in which to act. In describing the model,

Warner uses the example of chemotherapy for acute leukemia in children.

Before a technology is adopted or rejected it must be known. With regard to communication about technologies in the medical area, only the area of drugs has received the attention of researchers (120). Research on communication about drugs led to the description of a two-step model; information flows initially to physicians who are opinion leaders, and through informal channels, these leaders then transfer information to their followers (217).

The sources of information about technologies have been little studied. One study indicated that physicians specified drug companies' representatives as their most important source of information on new drugs (63). How the evaluations of technologies may affect their adoption has not been studied. It is clear, however, that the communication from researchers to practitioners is inadequate in both amount and quality.

A number of factors have been shown to influence the adoption of technologies. These include the characteristics of the technology, the complexity of understanding and using it, and the observability or visibility of its results (217). Characteristics of the adopter, including a cosmopolitan outlook have also been stressed (100). Large, complex, acute-care hospitals with medical school affiliations accept innovations more readily (176). Almost all the studies of adoption have focused on that of institutions like hospitals, and little is known about the adoption of technologies in practice situations.

Much research assumes physician dominance in decisionmaking (176). When there is concern about the slowness of change, physician conservatism is blamed. When premature adoption of technology is seen as the problem, physicians are considered to be uncritical and technology-hungry. Considerable homogeneity is assumed among physicians. Greer (101) has questioned these assumptions through research, still in progress, involving 362 focused interviews of those in the health care system, including 201 physicians. She found that community practitioners are general-

---

\*This section is based on Banta, Burns, and Behney, 1982 (9).

ly not interested in gaining influence in the hospital, and have little effect on the acquisition of technologies. Medical technologies were more often acquired through the actions of hospital administrators and hospital-based physicians than at the demands of patient-admitting community physicians.

From the standpoint of public policy, the key question is what characteristics of the medical environment affect adoption (96). These factors can be manipulated. They include financing methods, market conditions, and Government programs. The growth of third-party payment is without doubt related to the increasing use of medical technologies and increasing medical expenditures (167). The extent of coverage and methods of payment promote expensive hospital technologies and discourage preventive, rehabilitative, and ambulatory ones. Existing fee-for-service schedules reward the provider generously for diagnostic and curative services that rely on high technology. For example, a recent analysis in California showed that gastroscopy costs the physician \$40 to \$50, while Blue Shield pays up to \$240 for the procedure (205).

A key regulatory program influencing adoption is the drug regulation program of the Food and Drug Administration (FDA). FDA is required to approve all new drugs as efficacious and safe before they are marketed. In 1976, FDA authority was extended to medical devices (see ch. 3 for a fuller discussion of FDA regulation). FDA processes generally slow the adoption of technologies. A considerable body of research has shown that the licensing of drugs in the United States is relatively slower than in other countries and that the lag can in part be attributed to FDA (200). Since many technologies have diffused prematurely, however, it is not clear whether this delay is good or bad. Many other Federal and State programs directly or indirectly affect the adoption of medical technologies through regulation and financial means.

### The Use of Medical Technologies

While there are clearly some relations between adopting and using technology, they have not been clearly characterized. Some suggestive re-

search in this regard has shown that hospital beds tend to be used regardless of the health problems or demographic characteristics of an area population (191). The ready availability of laboratory tests through automation has apparently stimulated their rapid increase (227). Cromwell and his colleagues, however, report that nonprofit hospitals in Massachusetts use certain diagnostic equipment at only 50 to 60 percent of capacity.

A surprising finding is the highly variable relation between patient needs and technology use (195). This is true even in the case of specific technologies addressed to clearly defined medical conditions. Wennberg and Gittelsohn (249) found that rates of common surgical procedures vary greatly in small areas of New England, for example, even when the areas are contiguous and demographically similar.

Physicians' training and their role in society are important factors in technology use. The sociological literature on professionalism and on physician dominance is large. Physicians are professionals granted a high degree of autonomy (80). They are also agents of the patient who attempt to provide the best possible care, regardless of cost. Because the patients pay little or nothing for procedures directly, and they work in a system that rewards the use of technology with both profits and prestige, physicians have strong reasons to use technology (247). The development of medical specialties has also affected technology greatly. Specialties have developed in response to professional, technological, and economic interests in the past (212), and will most likely continue to respond to these interests. The United States is faced with a potential excess of physicians (228), who could respond to the resulting pressure by entering specialty practice and maintaining their incomes by using specialized technologies more intensively.

Malpractice suits apparently encourage the use of technologies like skull X-rays (15), electronic fetal monitoring (8), Cesarean sections (140), and clinical laboratory testing (202). The dynamic nature of malpractice has been little studied. An overemphasis on technology and a correspondingly diminished concern on the part of the physician can dehumanize medical practice. Such dehuman-

ization has been found to be associated with high rates of malpractice litigation (241).

Institutional factors affect technology use. The evaluations of prepaid group practices showing they led to fewer hospitalizations and less use of expensive technology were an important force in establishing health maintenance organizations in the 1970's. Similar evaluations are now encouraging the "competitive" strategy that is the latest policy. Such special medical institutions have been seen as a counterforce to the negative features of physician autonomy, but they also may diminish the physician's commitment to the interests of the patient and lead to a loss of the caring function of medicine (155). This could lead in turn to increased malpractice claims, a corresponding increase in technology use, and other problems.

As has already been stressed, fee-for-service payment to physicians and cost reimbursements to hospitals reward for providing more services. Existing fee scales reward more lucratively a physician's time using sophisticated technology than the physician's time in counseling (203). The spectacular rise in the use of ancillary services such as laboratory testing is related to specialization and extent of insurance, as well as payment methods. One study indicated that the greater use of nine such medical services accounted for about 40 percent of the increase in hospital operating costs from 1968 to 1971 (187).

The involvement of a profitmaking industry certainly affects the use of technology. The drug and device industries spend a large amount of money to promote their products. As mentioned previously, physicians say that the agents of drug firms are their most important sources of information about drugs.

### **Abandonment of Medical Technologies**

While researchers have been enthralled with the adoption of technologies, little has been done toward understanding their abandonment. McKinlay (150) describes a commonsense view of the "erosion and discreditation" of medical technologies. The initial enthusiasm for the technology when it was an innovation wanes and its applications are not so global as once thought. Sometimes

a scandal abruptly cuts short the life of a technology, thalidomide, for instance. More often, it is eclipsed by a new innovation. Finally, McKinlay says, "it is relegated to that great dust heap called History. "

In one of the very few attempts to analyze the abandonment process using empirical evidence, Finkelstein and Gilbert (72) examined the decline in use of eight drugs over the period 1964 to 1982. Seven had been introduced between 1963 and 1972, after the 1962 Amendments to the Food, Drug, and Cosmetic Act (see ch. 2) and one, tolbutamide (a hypoglycemic agent used by diabetics to lower blood sugar) which had been introduced earlier, but which experienced its decline during the later period.

Finkelstein and Gilbert began, for the sake of argument, with the assumption that abandonment would share features with adoption: that opinion leaders would first act on negative information about a drug, followed by the rest of the medical profession. Such a pattern represents the S-shaped curve. Their results suggest that, for the eight drugs studied, the pattern of abandonment does not fit the S-shaped curve. Declines in use were generally more precipitous, arguing that perhaps "physicians are sometimes affected directly by external information stimuli without the need for processing by an intermediary opinion leader. " Based on their findings, Finkelstein and Gilbert suggest that more investigations using empirical data could profitably be undertaken to systematically characterize alternative models for the abandonment and adoption of medical technology. The ultimate value might lie in better understanding of the influences on physicians in adopting and abandoning technologies.

### **RCTS and the Diffusion Process**

As the preceding sections have indicated, the reasons that medical technologies are adopted and used are far more complex than "simply" evaluating the evidence from RCTS and making reasonable decisions on that basis. The impacts of RCTS must be seen in this broader context, and efforts to increase their impact must consider the economic, regulatory, and institutional influences on adopting and using medical technologies.

The appearance of RCT results is not the start of a decisionmaking process about a medical practice, but comes after some diffusion has already taken place. Physicians may already have some personal experience with the technology, which may sway them in one direction or the other. RCTS are rarely conducted before new technologies are widely diffused (201). Banta and Thacker (8) document the widespread diffusion of electronic fetal monitoring despite the lack of evidence that it improves birth outcomes.

RCTS figure in two distinct processes: synthesis and consensus development. Synthesis is the process of integrating the findings from different studies and developing generalizations based on the results. All types of studies, both laboratory and clinical, may be considered in synthesis. Techniques for synthesis range from elementary qualitative procedures to sophisticated statistical manipulations.

The traditional approach to synthesizing research is the literature review. Typically, a reviewer selects a set of studies believed to be most relevant and summarizes the evidence. Because of the limitations inherent in literature reviews, efforts have been made to develop more systematic procedures to integrate and interpret sets of research evidence.

A simple structured synthesis technique involves organizing a body of literature according to a prespecified set of criteria and is actually a classification procedure (135). Sometimes called the “voting method,” this synthesis technique involves selecting a particular sample of evaluative studies of a technology, coding some aspect of the design and/or conceptual framework, classifying observed outcomes as to whether they are favorable, neutral, or unfavorable (i. e., “taking a vote”), and then constructing tables of research findings.

A rigorous statistical approach to research synthesis is a quantitative synthesis technique called

meta-analysis (93). This technique uses the actual results of studies and permits the determination, across a set of studies, of the magnitude of treatment impact. Meta-analyses are useful in assessing treatments for which a large number of studies are available and findings across studies seem to have great variability.

A number of organizations carry out synthesis activities. OTA reports have included a number of syntheses of specific technologies. Case studies prepared for *The Implications of Cost-Effectiveness Analysis of Medical Technology* (229) synthesize results of all types of research in their assessments. The activities of the former National Center for Health Care Technology and currently, the Office of Health Technology Assessment (National Center for Health Services Research, Department of Health and Human Services), are synthesis activities carried out by the Federal Government, in general with the aim of making statements about risks and benefits of technologies. In the private sector, the American College of Physicians and the Blue Cross and Blue Shield Association have specific programs of medical technology evaluation which use synthesis techniques.

Consensus development is a group decision process designed to produce a “consensus statement” about a medical technology, that can be accepted by clinicians, researchers, and the public. The statement should identify what is known and not known about the technology, in terms of the safety, efficacy, and appropriate conditions for use. The major sponsor for consensus development is the National Institutes of Health, through the Office of Medical Applications of Research. Unlike some of the structured synthesis techniques, consensus development conferences have no specific theoretical basis for their format. Consensus statements are widely distributed by NIH to the leading medical journals.